

Learning Genres by Using Song Lyrics

An NLP Project by Joshua Gottlieb

Problem Statement

Classifying music genres is a highly subjective subject and acquiring genre labels often requires expensive expert input.

Is This Actually Reasonable?

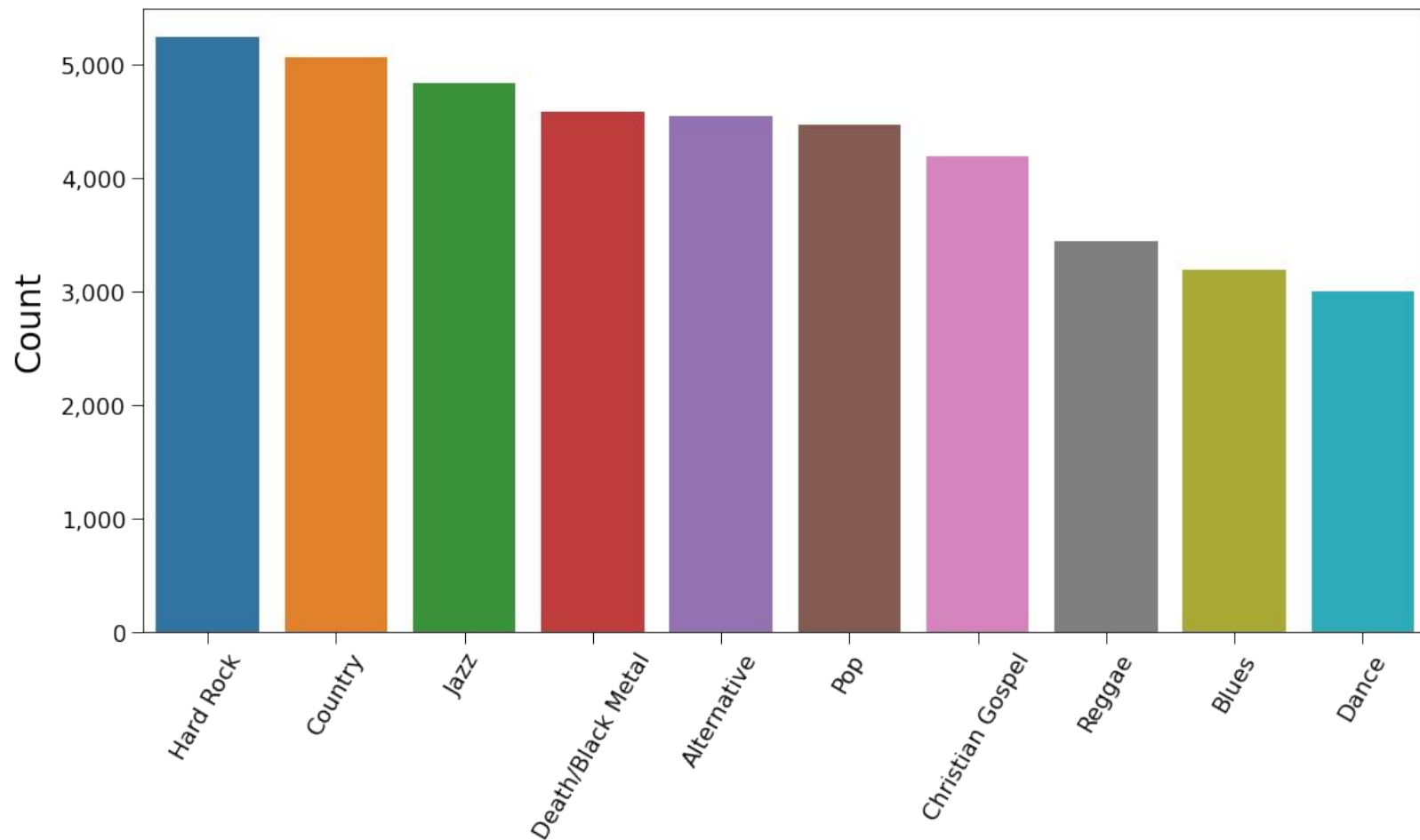
Short Answer

Not really.

Long Answer

Well, it depends...

Record Count for Each Genre Used



N-grams and Basic Word Context

1-Gram	2-Gram	3-Gram
The	The Margherita	The Margherita pizza
Margherita	Margherita pizza	Margherita pizza is
pizza	pizza is	pizza is not
is	is not	is not bad
not	not bad	not bad taste
bad	bad taste	
taste		

[From An Introduction to N-grams](#)

Topic Modeling - pyLDAvis

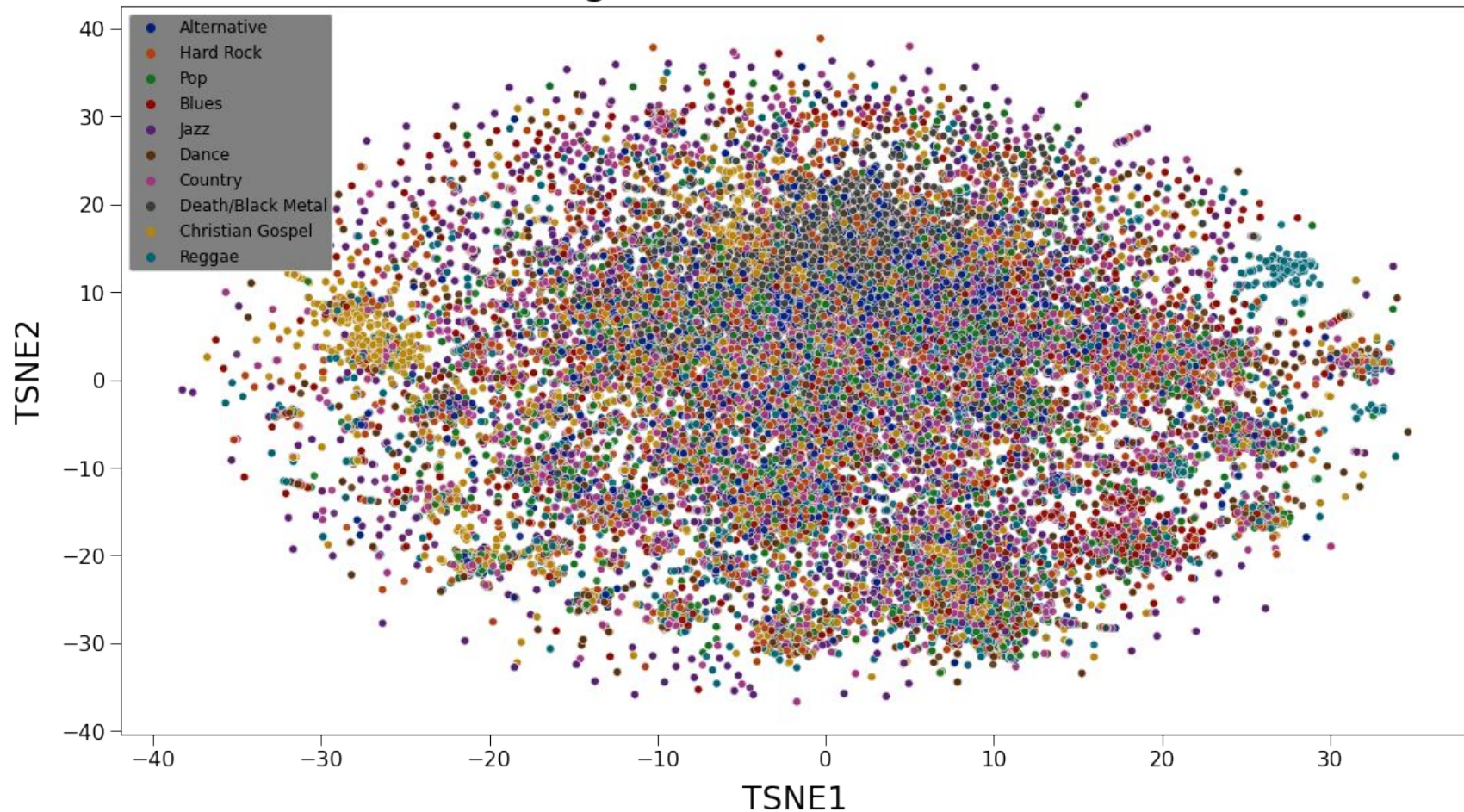
Highest Coherence Score - 8 Topics

(Open pyLDAvis topic modeling with 8 topics)

But we have 10 genres in our dataset?

(Open pyLDAvis topic modeling with 10 topics)

TSNE Gensim 4-gram Encoded Documents Train Set

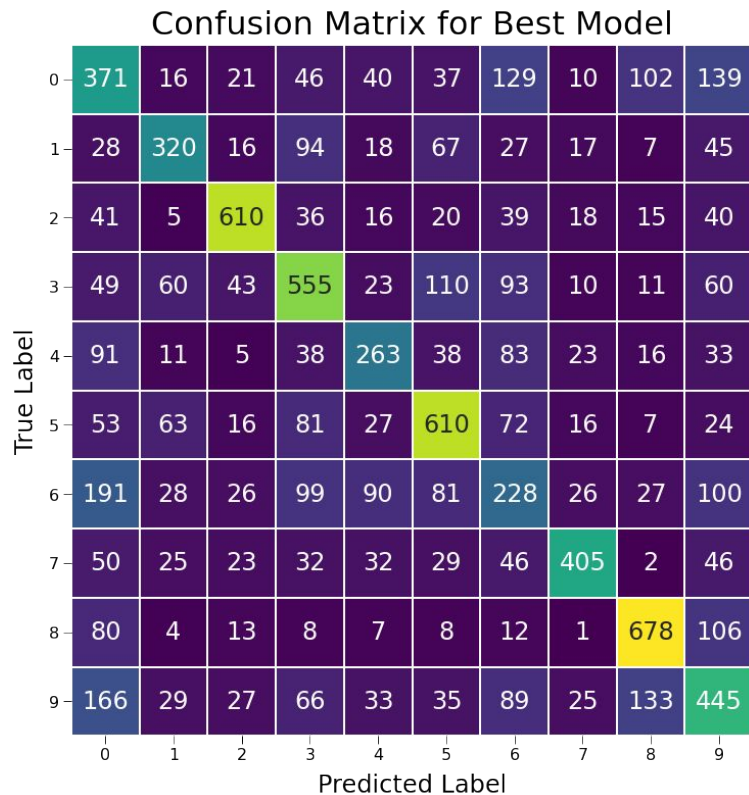


Models and Scores - Accuracy

	Count Vectorizer Unigrams	Count Vectorizer Bigrams	Gensim Bigrams	Gensim Trigrams	Gensim 4-grams	Averaged Glove	Averaged Word2Vec
Multinomial Naive Bayes	0.357	0.356	0.502	0.502	0.500	N/A	N/A
Untuned XGBoosted Tree	N/A	N/A	0.495	0.495	0.499	0.468	0.466
Tuned XGBoosted Tree *	N/A	N/A	0.523	0.525	0.526	0.463	0.478

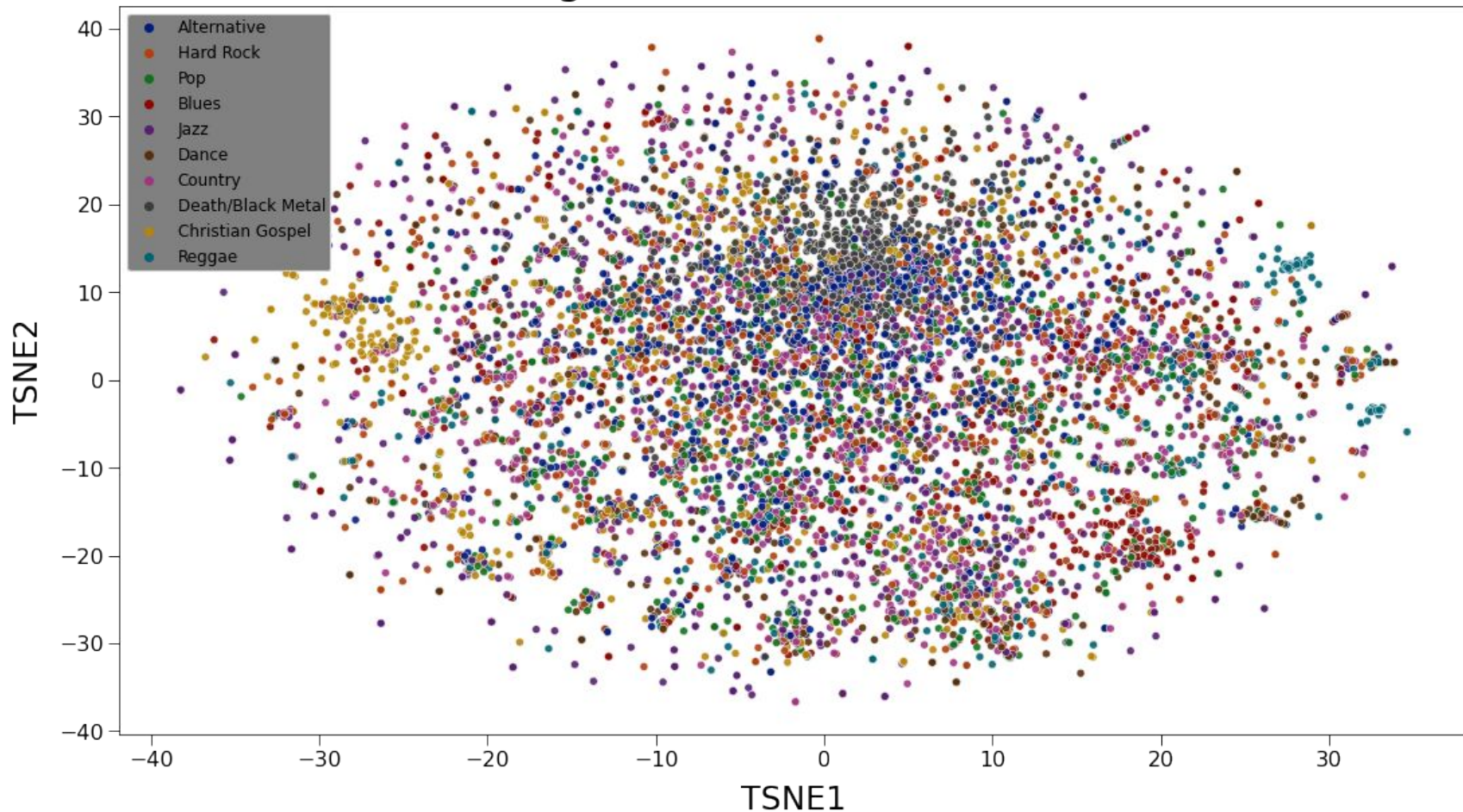
Best Model Confusion Matrix

0	Alternative
1	Blues
2	Christian Gospel
3	Country
4	Dance
5	Jazz
6	Pop
7	Reggae
8	Death Metal / Black Metal
9	Hard Rock



Accuracy = 0.526	Precision	Recall	f1 Score	Support
Alternative	0.33	0.41	0.37	911
Blues	0.57	0.5	0.53	639
Christian Gospel	0.76	0.73	0.74	840
Country	0.53	0.55	0.54	1014
Dance	0.48	0.44	0.46	601
Jazz	0.59	0.63	0.61	969
Pop	0.28	0.25	0.27	896
Reggae	0.74	0.59	0.65	690
Death Metal / Black Metal	0.68	0.74	0.71	917
Hard Rock	0.43	0.42	0.43	1048

TSNE Gensim 4-gram Encoded Documents Predictions



Conclusion

- Lyrics are not a very strong identifier of genre.
- However, some genres are quite distinct, while others are very generic.

Things to Test in the Future

- Drop or merge genres based off of topic model suggestions and domain knowledge.
- Find songs that can be "genre-swapped" based off of similarity in lyrics.

Thank you for listening!

Presenter: Joshua Gottlieb

Email: joshuadavidgottlieb@gmail.com

LinkedIn: [linkedin.com/in/joshua-gottlieb](https://www.linkedin.com/in/joshua-gottlieb)

GitHub: github.com/JoshuaGottlieb